NPS55-80-016

NAVAL POSTGRADUATE SCHOOL

Monterey, California

EXPERIMENTS WITH VOICE INPUT
FOR COMMAND AND CONTROL:
USING VOICE INPUT TO OPERATE
A DISTRIBUTED COMPUTER NETWORK

by

Gary K. Poock

April 1980

...ed for public release; distribution unlimited.

...ed for:

Naval Electronic Systems Command
Washington, D. C. 20360

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral J. J. Ekelund
Superintendent

J. R. Borsting
Provost

This investigation was sponsored by Mr. C. C. Stout, NAVELEX, Code 330 and Mr. W. J. Dejka, NOSC, Code 8302. The work was performed by the author at the Naval Postgraduate School, Monterey, California.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NPS55-80-016	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Experiments with Voice Input for Command and Control: Using Voice Input to Operate a Distributed Computer Network.		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Gary K. Poock		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, Ca. 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N0003980WR09041
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, CA. 93940		12. REPORT DATE April 1980
		13. NUMBER OF PAGES 48
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Naval Electronic Systems Command Washington, D. C. 20360		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

VTAG, Voice Recognition, Automatic Word Recognition

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes an experiment in which military officers used voice recognition equipment to verbally enter commands to a computer network similar to that of a command and control center, or a shipboard information center.

Results suggest that voice input can result in faster operation and fewer errors for such activities and give the operator more time to do other

OVER

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract Cont.

things. In addition, operators are very comfortable using voice input because of the more natural man-machine interface.

FOREWORD

This investigation was sponsored by Mr. C. C. Stout, NAVELEX, Code 330, and Mr. W. J. Dejka, NOSC, Code 8302. The work was performed by the author at the Naval Postgraduate School, Monterey, California.

This report is the first in a series concerned with the possible applications of using voice recognition technology in command and control tasks. A condensed version of this report was presented at the "Voice Interactive Systems: Applications and Payoffs" conference held 13-15 May, 1980 in Dallas, Texas.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXPERIMENTS WITH VOICE INPUT FOR COMMAND AND CONTROL: USING VOICE INPUT TO OPERATE A DISTRIBUTED COMPUTER NETWORK

I. EXECUTIVE SUMMARY

This paper describes an experiment in which military officers used voice recognition equipment to verbally enter commands to the ARPANET, a large distributed network of computers which are geographically located around the United States and other countries.

The objective was to determine if it was at all feasible to operate this network using commercially available state-of-the-art voice input equipment, and to compare this mode of entry with the normal manual typing input method.

Twenty-four military officers who already knew how to operate the ARPANET participated in the experiment. They were initially introduced to the voice equipment and then allowed to practice with it over a period of a few days until they felt "comfortable" with it. They had previously used the ARPANET for hundreds of hours using manual typing input so the amount of time they spent practicing with the voice equipment was a subjective feeling on their part as to when they were comfortable with it. The average subject practiced for 3.26 hours with the voice recognition equipment and then told the experimenter he/she was ready to participate in the experiment.

The experiment was then scheduled for an evening or weekend when the load average was under 3 on the host computers to insure fast network response times.

In the experiment, subjects followed a fixed scenario of instructions in which they accessed the ARPANET, logged into different host computers,

read messages, sent messages, checked for new mail, read files, transferred files between host computers, deleted files, and interconnected host computers. Each subject performed this scenario four times with either voice input first or typing input first, and then performed it four times with the other method of input. The scenario was designed to take about 10 minutes to perform, but the actual performance times ranged from 6 to 18 minutes. In order to measure any free time the subjects had while carrying out the scenario, a secondary task was included in which they transcribed information, by hand, from civil aviation weather reports onto a data sheet. Their main task therefore, was to run the ARPANET according to the scenario, but during any free time they were to transcribe the aviation weather data.

Keeping in mind that the average subject used the voice input method a little over 3 hours before doing the experiment, the results are quite significant.

The results, averaged across all trials of the experiment, show:

1) Voice input was 17.5% faster than manual typing input.

2) Manual typing input had 183.2% more entry errors.

3) Voice input allowed subjects to transcribe 25.0% more aviation weather information than during manual input.

These results are all statistically significant ($p < .05$) and suggest it is feasible to use current (1979) commercially available voice recognition equipment to run many standard operations of an ARPANET type network.

In an era when so much is said and written about declining productivity in America, voice input technology may be one solution to helping reverse this trend. We have observed here, that with minimal practice, the job was done 17.5% faster and at the same time, 25.0% more was done on another task.

What would happen if experienced voice input subjects were used?

II. INTRODUCTION

This paper describes an experiment in which a Threshold Technology, Inc., Model T600 discrete utterance voice recognition system was used to command the running and operation of the ARPANET.

III. OBJECTIVE

The objective of this experiment was to determine if it was at all feasible to operate a distributed computer network using voice input. The ARPANET was used in an unclassified mode to simulate the types of commands and operations used in and between military command centers. The ARPANET technology is the basis for the Advanced Command and Control Architectural Testbed (ACCAT) which is a classified subnet of the ARPANET on which several command centers are linked together for the purposes of testing and examining new software and hardware ideas applicable to command and control. Command centers on this network are located at installations such as the Naval Postgraduate School (NPS) in Monterey, California, the Naval Ocean Systems Center (NOSC) in San Diego, California, and CINCPACFLT in Hawaii.

Future voice input experiments will be run on this classified network in addition to the unclassified ARPANET.

IV. SUBJECTS

Twenty-four subjects participated on a volunteer basis with no monetary or other incentive. They included 23 male military officers from the Army, Navy, Air Force and Marine Corps, and one civilian female from the National Security Agency. Nineteen were enrolled in the Command and Control curriculum at NPS, 2 were enrolled in the intelligence curriculum at NPS, and 3 were military staff members at NPS. Experience levels in the military ranged from Lieutenant to Commander and from Captain to Lieutenant Colonel.

All subjects were experienced in using the ARPANET with manual typing input from a keyboard.

None of the subjects had ever used voice recognition equipment and only one had ever seen such equipment used.

3. INITIAL TRAINING AND EQUIPMENT USED

Subjects individually met with the experimenter initially and were given a subjective questionnaire regarding their opinions about using voice input versus manual typing input. At this time, they were also given a typing ability test.

They were then told about the basic ideas of how the voice recognition equipment worked so it could recognize what they would say and were also shown how we would be training the equipment for recognition.

The Model T600 Threshold Technology, Inc. voice recognition unit had several added memory modules which allowed up to 256 two-second voice utterances to be used. In this experiment, 180 of the possible 256 utterances (an utterance is any continuously spoken pattern of speech <u>up to</u> 2 seconds long, or as short as .1 of a second) were actually entered into the voice recognition unit although only about 75 utterances were actually needed in the experiment. The maximum length of two seconds for any utterance is a limitation imposed by the manufacturer.

The voice recognition unit also contained a magnetic tape cartridge unit which allowed the experimenter to record individual subject's voice patterns and ARPANET commands after the subject trained the machine initially. Then, when the subject came back to use the equipment at later times, the magnetic tape cartridge was simply read back into memory and the subject was ready to give voice input commands. (This is a nice feature as it allows one to take the equipment anywhere and connect to any computer or computer

network without relying on the host computer to store voice patterns. The tape cartridge feature also allows one to have a tape available for each type of task one might do. Then, if one switches to a new task which requires several hundred utterances unique to that task, one simply loads another tape cartridge containing the voice patterns and commands for that task.)

In this experiment, we also used the unbuffered mode which means that if the voice recognizer accepted a voice input, an ASCII character stream was immediately sent to the host computer without any verification by the operator that the voice recognizer had correctly interpreted the voice input. This allows for the possibility that one might say one thing but the voice recognizer "thinks" you said something else and therefore transmits the wrong ASCII stream. If an utterance is totally unacceptable, the voice recognizer just beeps. We could have guaranteed absolutely no input errors to the host computers if we had used the buffered mode which simply displays up to 128 utterances in series on a CRT and does not transmit the ASCII stream of characters until the operator verifies the stream and gives permission to transmit to the host computer.

In brief then, this voice recognition equipment allows for up to 256 utterances and with each utterance is associated an ASCII output stream. The subject can speak as many utterances as he wishes, as long as there is a .1-second delay between utterances. During an utterance, one must speak continuously for up to 2 seconds, and the voice recognizer then looks for at least a .1 second pause which is a signal to the recognizer that the old utterance has ended and a new utterance may be coming. Therefore, in normal talking, the following works fine if a .1 second pause is inserted where

indicated: "Select a map of the Med (pause) Show all Russian submarines (pause). How much fuel do they have? (pause) What is their destination? (pause)."

For this experiment, each subject trained the voice recognizer 10 times for each of the utterances and was then told he could practice running the ARPANET with voice commands. He could practice as much or little as he wanted during the next week until he felt comfortable using voice input. Then he was to tell the experimenter he was ready to do the actual experiment.

Subjects practiced from 1 to 8 hours with the average being 3.26 hours. This is important to keep in mind now that the results which follow are based on subjects who have used typing input to the ARPANET for hundreds of hours and have only used voice input for about 3 hours.

VI. EXPERIMENTAL PROCEDURE

The experiment was run in the evening or on weekends so the load average would be under 3 on the ARPANET hosts used. This, in fact, occurred with each of the 3 host computers used in the experiment. Two of the hosts were in southern California and one in Massachusetts. They were accessed from the NPS Terminal Interface Processor (TIP) located at NPS.

Based on the initial typing ability test, subjects were split into 2 groups called "SLOW" and "FAST" typers. The actual typing abilities ranged from 17 to 49 words per minute.

The actual experiment required subjects to follow a specific step-by-step scenario of instructions which required them to access the ARPANET, log into host computers, read messages, send messages, check for new mail, read files, transfer files between host computers, delete files, and interconnect host computers. The scenario was designed to take about 10 minutes to go through its steps one time. This scenario can be found in Appendix II.

The scenario was performed 4 times by each subject using voice input and 4 times using manual typing input. Half of the "SLOW" typers performed 4 trials through the scenario using typing input first, followed by 4 trials using voice input. The other half used voice input first followed by 4 trials using typing input. The "FAST" typing group was likewise counter-balanced with half using voice first and half using typing first.

A conceptual design for the experiment is shown in Figure 1. This is a three-factor nested design with repeated measures over trials. However, each subject is nested within only 1 of the typing ability conditions.

VII. SECONDARY TASK

In addition to performing the main task in the scenario set of instructions as fast and accurately as possible, subjects were given a stack of civil aviation weather reports with a blank data sheet for each report. When the subject had spare time between steps of the main scenario when the host computer might be transferring a file or something, the subject was to read the data sheet and record the appropriate data from the aviation weather report. For example, a data sheet might ask for runway visual range, fog conditions and cloud cover. Subject was to find the correct alpha-numeric information on the weather report and write it on the data sheet. When done with one data sheet, he proceeded to the next one as soon as possible. The data sheets did not always ask for the same information and the weather reports had random alpha-numeric information on them to prevent any pattern of learning.

After the experiment was finished for each subject, they were given the same questionnaire they had taken about two weeks before concerning their opinions and views on manual typing input and voice input.

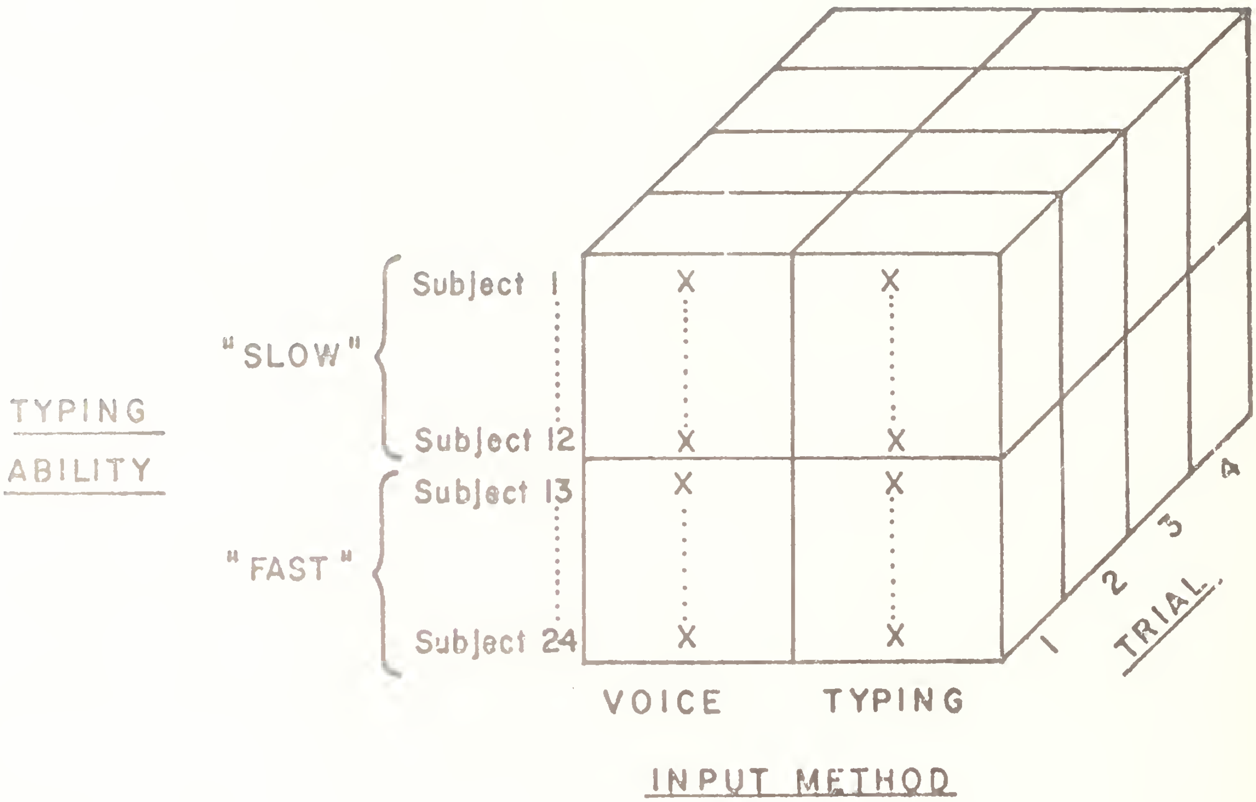

FIGURE 1. CONCEPTUAL DESIGN OF THE EXPERIMENT

VIII. DEPENDENT VARIABLES

During all trials, the following were measured:

1) Time to complete the scenario.

2) Number of input command errors to the computer network.

3) Number of characters transcribed correctly on the secondary task.

Note: We were interested in the number of times the network was instructed to do something wrong. Therefore, on typing input for example, if a command input was typed in wrong, it was counted as one error, whether there was one or several actual keystrokes typed wrong. Similarly, for voice input, if a subject spoke the wrong scenario command, the voice recognizer may have recognized the voice input correctly, but it would be a wrong command to the host and therefore was an error. Likewise, if the voice recognizer incorrectly identified a voice input and sent out the wrong command, this was an error. We were not interested in detailed analysis of how many times one voice utterance might get confused with another, i.e., the word "five" confused with the word "nine," etc.

In addition, we had ranked data from the subjects on their 'before and after' opinions on the questionnaire. The questions were ranked on a scale from 1 (strong feeling for manual input) to 7 (strong feeling for voice input) with 4 in the middle meaning neutral feeling between voice and typing input modes. These questions can be found in Appendix I.

IX. RESULTS

A. Results for Scenario Times

Figure 2 shows the times taken to perform the set of actions in the scenario. Table I shows the statistical results from the analysis of variance on times. (An α level of .05 had been chosen in the original experimental design. Therefore, when a result is discussed as being significant in this paper, it will mean that there is only a 5% chance or less that we are wrong when we say there was a significant difference in certain conditions.)

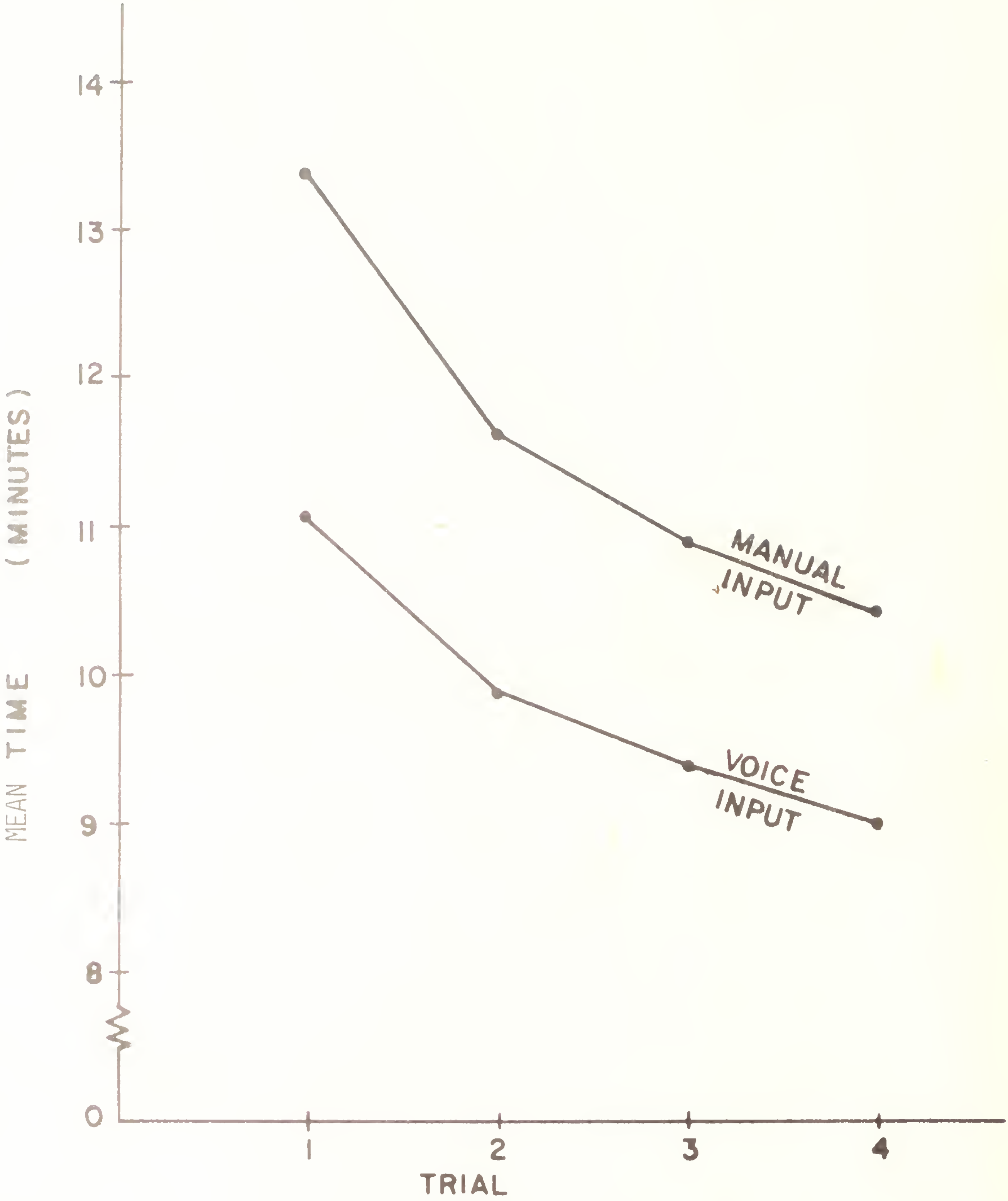

FIGURE 2. ELAPSED TIME TO PERFORM THE ENTIRE SCENARIO

TABLE I. Analysis of Variance for Scenario Times

Source	df	MS	F
Between subjects	23		
T (typing ability)	1	4.69	
Subj. w. groups	22	8.45	
Within subjects	168		
I (input method)	1	140.97	45.33*
T x I	1	4.21	1.35
I x subj. w. groups	22	3.11	
Tr (trials)	3	57.72	190.50*
T x Tr	3	.16	
Tr x subj. w. groups	66	.30	
I x Tr	3	2.09	2.72
T x I x Tr	3	.43	
I x Tr x subj. w. groups	66	.77	

* $p < .01$

As can be seen in Figure 2, voice input was consistently faster than manual typing input by an average 17.5%. This is a statistically significant difference in favor of voice input and even more important when we consider the subjects had only used voice input for about 3 hours in their entire life.

There was also a significant decrease in time over trials with both methods as indicated in Table I. A range test showed a significant improvement in time between each trial. We will never know if more trials would have improved performance even more. Four trials were initially chosen under each method of input, and as it turned out, the actual experimentation time for each subject was about 2 hours which left most of the subjects quite fatigued and mentally exhausted.

There was no difference in typing ability with respect to times. Both "slow" and "fast" typers could consistently perform better using voice input.

B. Results for Errors

Figure 3 illustrates the errors input to the system. The ANOVA results in Table II indicate a significant difference in typing ability and the "slow" and "fast" typers are therefore illustrated separately in Figure 3. Under both manual typing and voice input methods, the "fast" typers consistently made more errors than "slow" typers.

Under manual typing input, this was evident to the experimenter because "slow" typers were generally slow but quite precise in what they typed. However, "fast" typers would "go like hell" and thus cause a series of errors all at once. This personal characteristic of the "fast" typers appears to carry over into their performance using voice input also, since Figure 3 shows "fast" typers having consistently more errors than "slow" typers when using voice input also.

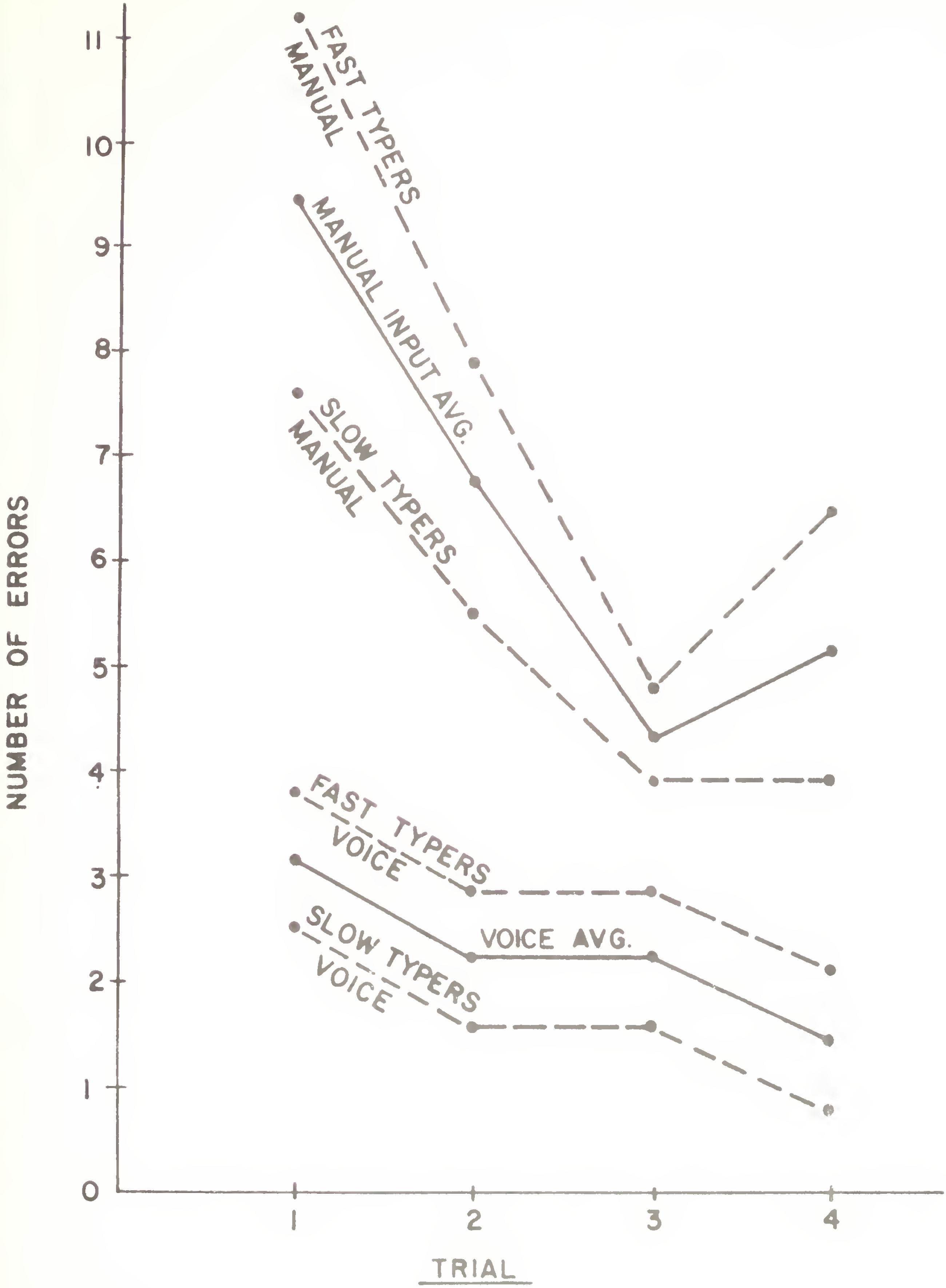

FIGURE 3. INPUT ERRORS TO THE COMPUTER NETWORK

TABLE II. Analysis of Variance for Errors

Source	df	MS	F
Between Subjects	23		
T (typing ability)	1	154.08	5.64**
Subj. w. groups	22	27.30	
Within Subjects	168		
I (input method)	1	825.02	64.51*
T x I	1	15.19	1.19
I x subj. w. groups	22	12.79	
Tr (trials)	3	96.33	13.11*
T x Tr	3	7.31	
Tr x subj. w. groups	66	7.35	
I x Tr	3	35.85	5.21*
T x I x Tr	3	.85	
I x Tr x subj. w. groups	66	6.88	

* $p < .01$ ** $p < .05$

Table II also shows an overall difference between voice input errors and manual typing input errors, as illustrated in Figure 3. Typing input averaged 83% more input command errors than did voice input.

One will recall from the previous section that there was no difference in scenario times for "slow" versus "fast" typers. If this is considered in combination with errors, it appears that any time improvement gained by "fast" typers is probably offset by their making more errors which requires more time for correcting input commands. Their scenario times are, therefore, similar to "slow" typers who don't do the scenario as fast, but also make fewer errors so spend less scenario time in correcting errors.

Table II also shows a significant difference in errors over trials. A range test indicated a significant decrease in errors from trial 1 to trial 2 to trial 3 over all conditions, but, on the average, trial 4 showed no improvement from trial 3. Table II also shows a significant interaction between trials and input method which is due mainly to the effect between trials 3 and 4 where errors increased under typing input but decreased under voice input. (See Appendix IV for voice recognizer performance details.)

C. Results for Secondary Task

Figure 4 shows the number of characters transcribed correctly on the secondary task using aviation weather report sheets. Since all subjects made so few errors on this task (five or less) the number of characters transcribed is actually the number correctly transcribed minus the number incorrectly transcribed.

Table III indicates a significant difference in input methods. These results are shown in Figure 4 illustrating 25.0% more information was transcribed on the secondary task during voice input than during manual typing

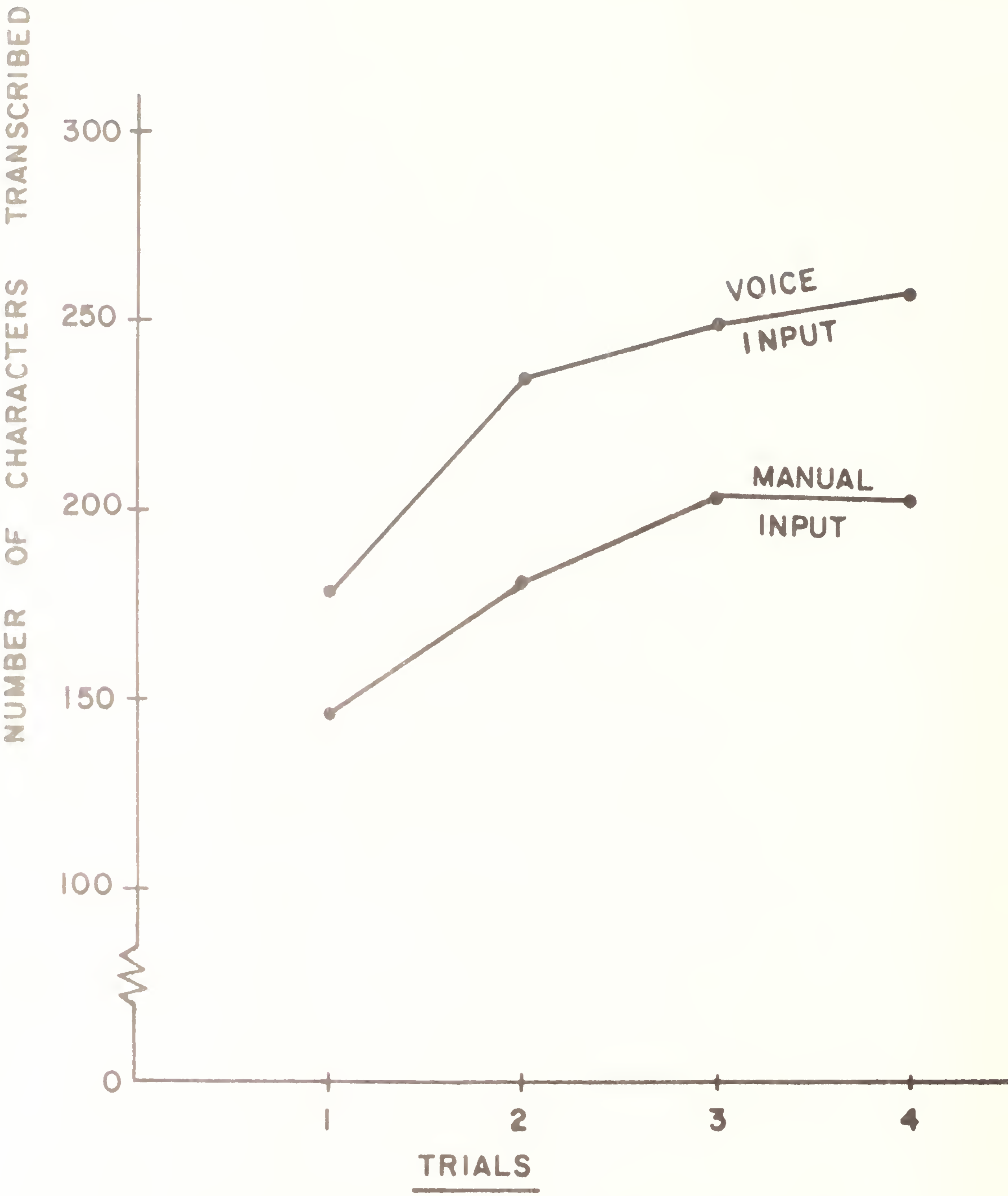

FIGURE 4. NUMBER OF CHARACTERS TRANSCRIBED ON THE SECONDARY TASK

TABLE III. Analysis of Variance for Characters Transcribed on the Secondary Task

Source	df	MS	F
Between Subjects	23		
T (typing ability)	1	30,451.69	1.68
Subj. w. groups	22	18,684.15	
Within Subjects	168		
I (input method)	1	101,292.19	24.32*
T x I	1	2,581.33	
I x Subj. w. groups	22	4,164.45	
Tr (trials)	3	46,599.58	59.86*
T x Tr	3	359.41	
Tr x Subj. w. groups	66	778.53	
I x Tr	3	1,137.69	1.09
T x I x Tr	3	913.72	
I x Tr x subj. w. groups	66	1,045.41	

* $p < .01$

input. In addition, there was a significant increase over trials. A range test showed significant increases in characters transcribed from trial 1 to trial 2 to trial 3, but no difference between trials 3 and 4.

D. Subjective Questionnaire Results

The subjective opinions received from each subject provided "before" and "after" data on the same questions. As described previously, these opinions were ranks on a scale from 1 to 7 and a nonparametric sign test (2 tailed; $\alpha = .10$) was therefore used to test for any general shifts in subjects' answers.

Subjects showed the following trends in their "before" and "after" feelings. The numbers following each item show the average response before and after, where the response scale was 1 for strong typing input feeling, 4 was a neutral feeling and 7 was a strong feeling for voice.

a) Subjects showed a significant shift in opinion concerning ease of input. Before the experiment, they had a feeling voice would be easier than manual input of commands to the computer, and after they felt even more strongly that this was the case (avg. before = 4.58; avg. after = 6.13).

b) With respect to whether they would be more frustrated using manual typing or voice input, subjects started out feeling manual typing would be more frustrating and felt even stronger about this after (avg. before = 3.42; avg. after = 2.63).

c) After the experiment, subjects felt more strongly that voice input allowed one more time and freedom to do other things than did manual typing input (avg. before = 5.88; avg. after = 6.63).

d) Subjects also started out with a feeling that voice input might allow more flexibility in entering items to a computer and after felt even stronger about this (avg. before = 3.92; avg. after = 4.58). This author had thought

they would think manual input was more flexible. However, since the vocabulary of utterances for each subject included all the single digits and the entire military alphabet, they actually had a lot of flexibility with voice also. For example, to "Forward" a message, the message system required an "F" to be input. They could simply say "Forward message" which would transmit the "F", but many of them also used "Foxtrot" of the military alphabet which also transmitted an "F."

e) When asked if they would be more relaxed using manual typing or voice input, their response showed no statistical change. They started out feeling they would be more relaxed with voice input and their feeling remained that way, (avg. before = 5.00; avg. after = 5.67).

Four questions were based on a scale from 1 to 7 with 1 meaning absolutely not, 4 meaning neutral and 7 meaning absolutely yes. These results were:

a) Subjects showed a significant change when asked if, in general, they liked the idea of voice input (avg. before = 6.00; avg. after = 6.50). They thought they would like it before and subsequently did.

b) When asked if they would like to use voice input in everyday tasks, if it were applicable, they showed a similar significant change (avg. before = 6.00; avg. after = 6.54).

c) When asked if voice input could be applicable in command and control tasks, subjects started out feeling quite positive and felt more strongly about this after the experiment (avg. before = 6.04; avg. after = 6.38).

d) When asked if voice input could be used in military tasks other than command and control, they felt before the experiment that it could and retained this opinion after (avg. before = 6.00; avg. after = 6.29).

Finally, the question "Does voice input provide a better man-machine interface?" was asked only at the end of the experiment. On the same "absolutely no" to "absolutely yes" seven-point scale, the average subject response was toward "yes" with an average response of 5.80.

X. OTHER OBSERVATIONS

A. There was no correlation between the amount of practice time subjects spent in becoming familiar with the voice recognition method and the fastest time in which they were able to perform the scenario using voice input. Likewise, there was no correlation between practice time and errors entered to the network.

B. Voice input offers a better man-machine interface because the user can operate under conditions familiar for him. In the current experiment for example, a carriage return was required quite often. Each user could use the voice command most comfortable for him, and in the case of carriage return, some subjects used "return," "carriage return," or "go" while others chose "do it," "send it," or "roger." In a few cases, a subject even requested that he be able to use two different utterances which sent out the same ASCII stream of characters, so if he forgot one of the utterances during the stress of performing the experiment, he could use his alternate command just as easily.

C. Voice input appears to reduce the problems of entering complicated strings of characters also. If a user needs to enter "*/(LEN=)*" he may make numerous mistakes in a manual keyboard entry mode, but with voice input, he can simply choose a phrase he likes to use and the above output ASCII stream is always the same and entered for him automatically.

D. Several subjects mentioned that with voice input they felt they had better command of the situation because they could see what the network was doing and at the same time their hands were free. With manual input they felt they were more at the mercy of the keyboard and concentrated more on typing the right characters rather than observing the big picture of what was going on.

E. Our particular models of voice recognition equipment contain a structuring feature which allows one to operate on a subset of the total 256 utterances. By only operating on a subset of the utterances, one would get faster recognition times. However, it is this writer's experience that structuring is not needed. Even when using all 256 possible utterances in the memory of the voice recognizer, the response time is so fast that it is practically impossible for the user to notice any delay. We commonly use all 256 utterances, and in such cases, we can enter a voice command to the recognizer, and before one can blink an eye, a host computer hundreds of miles away is replying. Therefore we currently find it not necessary to use structuring of any kind, although it is a topic for future research.

F. It is interesting to observe a behavioral phenomena when introducing people to voice input also. This author often gives demonstrations of various software products in the NPS command center. I can literally make many mistakes in manual typing when running a particular demo, and people will accept my poor typing ability and be happy. However, when I use voice input, I might make one mistake

an hour, but the observers will immediately notice it and say something to the effect that voice input is nice, but is not perfect and has a way to go. That is true, but it is interesting to note that moments before I made all sorts of manual typing input mistakes and it did not bother them!

G. We also found the portability of our units to be a nice feature. Since we do not depend on any foreign host computer to store the voice patterns, we can go anywhere with our units and be operational immediately.

XI. CONCLUSIONS

Based on the results of this experiment, 24 military officers were able to effectively operate a distributed computer network with minimal voice training. Considering that they already knew how to operate the network in a manual typing input mode, they were still able to operate the network faster using voice input, they made far fewer input errors with voice, and at the same time, managed to get 25% more work done on another task when using voice input than when using manual typing input.

The results suggest that voice input may be a technology which can be of benefit in command center operations, combat information centers and similar installations.

Future and/or current plans for our experiments include examining:

1) The use of voice input with military decision aids.

2) The use of voice input with interactive graphics.

3) The use of voice input by users during tactical computer games.

4) The use of voice input for human image interpreters

5) The use of voice input in NATO type command centers where multi-lingual users are prevalent. Pilot experiments have indicated that for the 10 training passes used for each utterance, we can enter 5 passes in English and 5 in German for a given utterance, and then the voice recognizer still appears to work quite well whether one speaks in English or German. If in fact we can make this work satisfactorily, we can effectively double the possible utterances from 256 to 512.

6) The effect of shipboard and command center environmental noises and disturbances on voice input.

7) The effect of multi-task mental loading on an operator and his voice input performance.

8) The amount of training required for effective use in various tasks.

BIBLIOGRAPHY

(This is a general list. Each of these contains many more references and/or papers on the subject.)

CONNOLLY, D. W. Voice Data Entry in Air Traffic Control. FAA, NAFEC report FAA-NA-79-20, August 1979.

LEA, W. A. and SHOUP, J. E. Review of the ARPA SUR Project and Survey of Current Technology in Speech Understanding. Speech Communications Research Lab, 606 West Adams Blvd., Los Angeles, California 90007, January 16, 1979.

SKRIVER, C. Vocal and Manual Response Modes: Comparison Using a Time-Sharing Paradigm. Naval Air Development Center, Report DTR 79-1, Warminster, Pa., January 1979.

_____, Proceedings of the Workshop in Voice Technology for Interactive Real-Time Command/Control System Application (R. Breaux, M. Curran, E. Huff, Editors). NASA Ames Research Center, December 6-8, 1977.

APPENDIX I

(Subjects were asked the following questions both before and after the experiment. Items 1 and 2 were yes or no responses. For Items 3 through 7, subjects marked their choice on a scale from 1 to 7 where 1 was a very strong feeling for manual input, 4 was marked neutral feeling, and 7 was a very strong feeling for voice input. Verbs were changed appropriately for questions when asked after the experiment.)

1. Have you used voice input before?
2. Have you seen voice input used before?
3. Which might be easier, manual typing input or voice input for communicating to a computer?
4. Would you be more relaxed using manual typing input or voice input?
5. Would you have more flexibility in entering items to a computer with voice input or manual typing input?
6. Would voice input or manual typing allow you more time and freedom to do other things?
7. Would you be more frustrated using voice input or manual typing?

(On Items 8-11, subjects marked their choice on a scale from 1 to 7, where 1 was "absolutely NO," 7 was "absolutely YES," and 4 was a neutral feeling.)

8. In general, do you like the idea of voice input?
9. In general, do you think you would like to use voice input in every day tasks yourself if it were applicable?
10. In general, do you think voice input would be useful for application in command and control tasks.
11. In general, do you think voice input could be used in military tasks other than command and control?

APPENDIX II

SCENARIO INSTRUCTIONS

1. GO TO HOST ISIE (host 116)
2. See if there is MAIL for EXPERIMENTAL
3. LOG INTO EXPERIMENTAL
   a) GET THE LOAD AVERAGE
   b) go into MSG
   c) Read the 1st message
   d) FORWARD the 2nd message to Poock
   e) Call the message "VOICE DEMO"

      NO CC:

      Don't add any new text

      Send it
   f) Exit to EXEC LEVEL
   g) Get the LOAD AVERAGE
4. TELNET TO ISIC
5. See if there is MAIL for C3DEMO
6. LOG IN TO C3DEMO
   a) List all the directory files
   b) Type out the file beginning with a Z
   c) Go into MSG
   d) Read the 3rd message
   e) Exit back to EXEC LEVEL
7. LOGOUT
8. DISCONNECT AND QUIT BACK TO EXEC LEVEL AT ISIE.
9. GET THE LOAD AVERAGE

10. FTP to ISIC

    a) Log into C3DEMO

    b) List the C3DEMO Directory on your TTY

    c) Get the remote file "LADDER.RUNFIL" to your local file "VOICE.RUNFIL"

    d) Break the FTP connection and Disconnect and Quit.

11. You are back at ISIE now

    a) Delete the file "VOICE.RUNFIL"

    b) Go into MSG

    c) Send a message as follows:

       TO: POOCK

       CC: C3DEMO

       SUBJECT: Pacific Report

       MESSAGE: All Units Ready

       WX report = clear

    d) Send it

    e) Exit back to EXEC LEVEL

12. Get the Load Average of the system

13. TELNET to BBNA

    a) Log in as NPS

    b) List all the directory files

14. LOGOUT of BBNA

    a) Disconnect and Quit Back to EXEC LEVEL at ISIE

15. Get the LOAD AVERAGE

16. LOGOUT.

APPENDIX III

Decoding Aviation Weather Reports from Civil Stations

U.S. DEPARTMENT OF COMMERCE
ENVIRONMENTAL SCIENCE SERVICES ADMINISTRATION
WEATHER BUREAU
SILVER SPRING, MD. 20910

DECODING AVIATION WEATHER REPORTS
Based on Instructions in Federal Meteorological Handbook No. 1, Surface Observations

STANDARD AVIATION REPORT FORMAT FOR MANNED STATIONS

EXAMPLE OF AN OBSERVATION AS FOUND ON HOURLY SEQUENCES

BC198	REMARKS. Visibility variable between 1/2 and 1 mile. REMARKS Ceiling variable between 900 to 1200 feet.
T2Ø	BASES AND TOPS OF CLOUDS Tops broken layer 2700 ft. msl. Height of bases not visible at the station precede sky cover symbol. "U" indicates layer amount unknown. If the report is more than 15 minutes old, the time (GMT) precedes the entry.
	REMARKS: Fog and Smoke hiding 3/10 of sky.
L11VR8	RUNWAY VISUAL RANGE: Runway 10L, Visual Range variable between 2600 and 5500 ft. in past 10 minutes. When visual range is constant for past 10 minutes, only the constant value is reported, e.g., R10LVR60+.
1Ø12	ALTIMETER SETTING: 29.67 inches. Three figures, representing units, tenths and hundredths of inches, indicate the altimeter setting. "Low" is used preceding figures to indicate values below 29.00 inches.
52E	WIND: 270° true, 13 kts. To decode direction, multiply first 2 digits by 10. If product is >500, subtract 500 and add 100 to speed. Gusts and squalls are indicated by "G" or "Q" following speed and peak speed following the letter.
31	TEMPERATURE: 66°F. A minus sign indicates temperatures below zero.
24	SEA LEVEL PRESSURE: 1014.6 millibars. Only the tens, units and tenths digits are reported.
26	WEATHER AND OBSTRUCTIONS TO VISION: Light Drizzle, Fog & Smoke. Symbols used in reporting weather and obstructions to vision are in Table 1. Algebraic signs (Table 1) following symbols indicate intensity.
7	PREVAILING VISIBILITY: Seven eighths statute mile and variable by the amount given in REMARKS.
2T4M1	SKY & CEILING Partly obscured sky, ceiling measured 1100 ft., variable broken, 3800 ft. overcast. Figures are height of each layer in 100s of feet above ground. A number preceding an X indicates vertical visibility into phenomena. A "V" indicates height varying by amount given in REMARKS. Symbol after height is amount of sky cover (Table 2). The letter preceding height indicates that height to be the ceiling and the method used to determine the height (Table 3).
	TYPE OF REPORT (Table 4): "R" omitted when observation is in hourly sequence.
	STATION IDENTIFICATION: Identifies report for Pittsburgh by using FAA identifier.

<u>Sample</u>

AVIATION WEATHER REPORT

<u>DATA SHEET</u>

REMARKS	____________
BASES AND TOPS OF CLOUDS	____________
RUNWAY VISUAL RANGE	____________
WIND	____________
TEMPERATURE	____________
PREVAILING VISIBILITY	____________
SKY AND CEILING	____________
STATION IDENTIFICATION	____________

(NOTE: A sample aviation weather report is shown on the previous page. A data sheet shown above was attached to each weather report. In the above case, subject would look for remarks on the report, copy down BC198 on the data sheet and proceed to the next item. The values on the aviation weather reports were all different and the items asked for on the data sheet were mixed up, i.e., sometimes a data sheet asked for WIND and other times not. When a weather report was done, subject went on to the next weather report and data sheet.)

APPENDIX IV

VOICE RECOGNIZER PERFORMANCE

DETAILS IN OPERATIONAL EXPERIMENT

Figure 3 in the text discusses input errors to the network. Although that Figure translates into a 3% error rate for voice, the data below show that actual performance of the recognizer in various categories. (If you say an utterance and the T600 does not recognize the utterance, then the T600 beeps and no ASCII output string is sent.)

TOTAL UTTERANCES in this operational experiment were 7,200 (i.e. 75 utterances per trial × 4 voice trials × 24 subjects).

Recognizer Details:

Category		% of time
1.	Correct Utterance AND Correct Output	96.80
2.	Correct Utterance AND Wrong Output	.76
3.	Correct Utterance AND No Output (Beep)	.36
4.	Invalid Utterance AND No Output (Beep)	.78
5.	Invalid Utterance AND Recognizer Put Out Something When it Should Have Beeped	1.30

Items in 5 above were caused mostly by the inexperienced subjects mumbling and trying to figure out where they were under the time pressure of the scenario and the secondary task.

Although the total error rate for the recognizer is about 3% and that shown in Figure 3 is about 3%, one should note Figure 3 is input errors to the network. Therefore Figure 3 is based on the errors in Category 2 and Category 5 plus operational input errors, where the recognizer worked correctly, but the subject entered the wrong command to the network when a different command was required by the scenario.

APPENDIX V

SUGGESTED VOCABULARY

The following phrases were suggested but subjects could use their own phrase instead if they wished. The 180 utterance vocabulary was entirely open with no branching to subsets of words during the experiment. The first one for example, GO TO ECHO, was 3 words spoken continuously to make one utterance, and similarly for the other phrases.

GO TO ECHO	C3 DEMO
CONTROL ALPHA	DELETE MESSAGE
GET FILE	DELETE FILE
VOICE RUNFILE	TYPE FILE
LADDER RUNFILE	BACKSPACE
FORWARD MESSAGE	SPACE
VOICE DEMO	CONTROL N
STRAIT OF HORMUZ	LOAD AVERAGE
AIR ROUTES	GARY POOCK
RUSSIAN VERSION OF HORMUZ	PACIFIC REPORT
CLOSE OUT CHARLIE	ALL UNITS READY
GENISCO ZERO PARAMETERS	DIRECTORY
THREE MAPS	TTY
LEVEL TWO VIEWER	ESCAPE
MEDITERRANEAN MAP	WEATHER REPORT
NORTH ATLANTIC MAP	REDSPHERE
SOUTH ATLANTIC MAP	CONNECT TO CHARLIE
SMILE	CHANGE DIRECTORY
QUIT	COLOR BLOCK
TYPE MESSAGE	CONTROL QUEBEC
HEADERS	TENEX
SEND MESSAGE	LOGIN NPS
GO	LOGIN NPS ONE
DASH	TELNET TO UNIX
COMMA	TELNET TO TENEX

TELNET TO TOPS2Ø
CONTROL HOTEL
COLOR BLOCK RUNFILE
EXAMPLE RUNFILE
CHANGE DIRECTORY TO POOCK
SEARCH AND RESCUE
LOAD GLD3
LOAD THE SERVER
LOAD THE GANN
CONNECT TO ECHO
CONTROL DELTA
INFO MAIL EXPERIMENTAL
MAIL CHECK C3DEMO
TELNET
DISCONNECT
FTP
GOODBYE
ISI CHARLIE
BBN ALPHA
ISI ECHO
MSG
EXIT
CONTROL O
CONTROL Z
CONTROL CHARLIE
LOGOUT
LOGIN C3 DEMO
LOGIN EXPERIMENTAL
C2 NET CONTROL PASSWORD
LOGIN C2 NET CONTROL
GO TO BBN ALPHA
LARRY SHACKLETON
MAIL BOX
GO TO SRI DASH KL
SEMICOLON
PERIOD
FROM
ASTERISK
UNDELETE FILE
TEXT EDITOR
MAIL STAT
STAMMER 2 RUNFILE
CONTROL BRAVO
CONTINUE
JACK WOZENCRAFT
REX STOUT
JACK DIETZLER
SEND FILE
ISI ALPHA
POOCK NPS PASSWORD
ACCAT BOX
LOGIN ACCAT BOX
ANSWER MESSAGE
FORWARD MESSAGE
NORTH
EAST
SOUTH
WEST
MY POSITION IS
GO TO CHARLIE
INFO DISK
DISK STATUS
WHOIS
LOGIN XCNO
WHARTON
SLANT
RECENT MESSAGES
NOT EXAMINED
LOGIN HOLLISTER
R SCHLAFF

KNELMS
SEND
TRANSMIT
MAIL CHECK STOUT
AT
ERASE
CANCEL
CLOSE CONNECTION
DOWN
CONTROL TANGO
SPHERE
UP IN DETAIL
DOWN IN DETAIL
ACCAT TITLE
MOVE IT DOWN
MOVE IT UP
MOVE IT LEFT
MOVE IT RIGHT
BREAK
SPIROGRAPH
USE THAT ONE
LEVEL TWO
GRAPHICS
MARBLES

PLUS the 10 digits and the 26 word military alphabet were also included.

DISTRIBUTION LIST

NO. OF COPIES

Defense Technical Information Center 12
Cameron Station
Alexandria, VA 22314

Library, Code 0142 2
Naval Postgraduate School
Monterey, CA 93940

Library, Code 55 1
Naval Postgraduate School
Monterey, CA 93940

Office of Research Administration 1
Code 012A
Naval Postgraduate School
Monterey, CA 93940

Dean of Research, Code 012A 1
Naval Postgraduate School
Monterey, CA. 93940

Naval Postgraduate School
Monterey, CA. 93940
Attn: J. Arima, Code 54Aa
R. Elster, Code 54Ea 1
D. Barr, Code 55Bn 1
W. Moroney, Code 55Mp 1
D. Neil, Code 55Ni 1
G. K. Poock, Code 55Pk 50
R. J. Stampfel, Code 55 1

E. L. Wiener 1
239-3
NASA
Moffett Field, CA 94035

Anthony Bessacini, Code 3522 1
NUSC
Newport, RI 02840

	NO. OF COPIES
Bruno Beek Rome Air Development Ctr/IRAA Griffiss AFB, Rome, NY 13441	1
John F. Boehm DIRNSA Attn: R542 Ft. Geo. G. Meade, MD 20755	1
R. Breaux Code N-711, NAVTRAEQUIPCEN Orlando, FL 32813	1
CDR Paul R. Chatelier OUSDR E (Research & Advanced Tech) Mil. Asst. TR & Pers Tech Rm 3D129, Pentagon Washington, DC 20301	1
Ralph L. Cleveland NFMSO, Code 9333 Mechanicsburg, PA 17055	1
Clay Coler Mail Stop 23902, NASA-Ames Research Center Moffett Field, CA 94035	1
Dianne Davis Naval Underwater Systems Ctr Code 3522 Newport, RI 02840	1
Edward DeGregorio Naval Underwater Systems Ctr Code 3522 Newport, RI 02840	1
Tice DeYoung U. S. Army Engineer Topographic Lab Research Institute Ft. Belvoir, VA 22060	1
Joe Dickinson U. S. Army Applied Tech Lab Ft. Eustis, VA 23662	1
William E. Gibbons Naval Air Development Ctr Warminster, PA 18974	1

No. of Copies

Harold C. Glass U. S. Postal Lab 11711 Parklawn Dr., Rockville, MD 29852	1
Henry W. Halff Code 458 Office of Naval Research Arlington, VA 22217	1
LT Steven D. Harris 6021 NADC Warminster, PA 18974	1
Warren G. Lewis Code 8231 Navy Ocean Sys Ctr San Diego, CA 92152	1
CWO2 Robert R. Lynchard ATC Training Dept. NATTC Memphis Millington, TN 38054	1
John T. Masterson U. S. Postal Lab Rockville, MD 20852	1
Don McKechnie AFAMRL Wright-Patterson AFB, OH 45433	1
Thomas J. Moore AFAMRL/BBA Wright-Patterson AFB, OH 45433	1
CAPT Vince Mortimer AFAMRL/BBM Wright-Patterson AFB, OH 45433	1
James D. Mosko Acoustical Sciences Deiv. Naval Aerospace Med Res. Lab NAS, Pensacola, FL 32504	1
CAPT Leslie K. Scofield Directorate of Training U. S. Army Signal Center Ft. Gordon, GA 30905	1

No. of Copies

C. Skriver 6021 NADC Warminster, PA 18974	1
S. Nils Straatveit Naval Underwater Systems Ctr Code 317 New London, CT 06320	1
Leahmond Tyre FLTMATSUPPOFFICE, Code 9333 Mechanicsburg, PA 17055	1
Eric Werkowitz AFFDL/FGR Wright-Patterson AFB, OH 45433	1
T. Weiner Code 4043 NADC Warminster, PA 18974	1
LT Jeff Woodard Rome Air Development Ctr/IRAA Griffiss AFB, Rome, NY 13441	1
Irv Alderman Army Research Institute 5001 Eisenhower Ave., Alexandria, VA 22333	1
John Phillips Code 7232 NOSC San Diego, CA 92152	1
Wayne Lea 889 Sanford Court Santa Barbara, CA 93111	1
Mr. C. C. Stout NAVELEX, Code 330 2511 Jefferson Davis Highway Arlington, VA 20360	1
Bruno Witte Code 8302, NOSC San Diego, CA 92152	1

No. of Copies

W. Dejka Code 8302 NOSC San Diego, CA 92152	1
LCDR J. Dietzler ARPA-IPTO 1400 Wilson Blvd Arlington, VA 22209	1
CAPT Mike Hayes ARPA-CYBERNETICS OFFICE 1400 Wilson Blvd Arlington, VA 22209	1
Richard Pew BBN 50 Moulton St. Cambridge, MA 20138	1
Tom Imperato 5525 Wilshire Blvd., Suite 205 Los Angeles, CA 90036	1
Marlin Thomas I. E. Dept. 113 Electrical Eng. Bldg University of Missouri-Columbia Columbia, MO 64211	1
M. Tolcott Office of Naval Research, Code 455 800 N. Quincy St. Arlington, VA 22217	1
G. Malecki Office of Naval Research, Code 455 800 N. Quincy St. Arlington, VA 22217	1
N. Greenfeld BBN 50 Moulton St. Cambridge, MA 02138	1
R. Bisbey USC-ISI 4676 Admiralty Way Marina del Rey, CA 90291	1

No. of Copies

R. Kolb NOSC, Code 824 San Diego, CA 92152	1
CDR R. Meinhold CINCPACFLT Box 6 Pearl Harbor, HI 96860	1
Gary Martins RAND Corporation 1700 Main St., Santa Monica, CA 90406	1
Earl Sacerdoti SRI, AI Center 33 Ravenswood Ave., Menlo Park, CA 94025	1
Daniel Sagalowicz SRI, AI Center 33 Ravenswood Ave., Menlo Park, CA 94025	1
Rick Butterworth SEI 1340 Munras Monterey, Ca. 93940	1
Russ Hammond SAI 1911 N. Fort Myer Drive, Suite 1200 Arlington, VA 22209	1
Don Connolly FAA NAFEC Atlantic City, NJ 08405	1
Stu Parsons 19740 Via Escuela Dr. Saratoga, CA 95070	1
Don Chaffin Industrial Engineering Dept. University of Michigan Ann Arbor, MI 48104	1
Walt Hancock Industrial Engineering Dept. University of Michigan Ann Arbor, MI 48104	1

	No. of Copies
D. McCall NOSC Code 8242 San Diego, CA 92152	1
Marvin Denicoff Office of Naval Research, Code 437 800 N. Quincy St. Arlington, Va 22217	1
John Schill NOSC Code 8123 San Diego, CA 92152	1
H. Morgan Wharton School University of Pennsylvania Room W-83, Dietrich Hall Philadelphia, PA 19104	1
Don Lesemann NOSC Code 7103 San Diego, CA 92152	1
Dan Schutzer NAVELEX, PME-108 Washington, D.C. 20360	1
H. G. Miller NAVELEX, PME-108-3 Washington, D.C. 20360	1
R. Kahn ARPA-IPTO 1400 Wilson Blvd Arlington, VA 22209	1